# Development and application of the statistical ecology in China

Zhang Wanli (张万里)

Open Research Laboratory of Forest Plant Ecology, Northeast Forestry University, Harbin 150040, P. R. China

**Abstract** The statistical ecology develops rapidly with the advancement of analyzing technology and method. This paper made a review about statistical ecology in China in following aspects: data collection, spatial and time patterns analysis, species-abundance and species-area relations, population dynamic, species affinity, community similarity and community cluster analysis, and community ordination.

**Key words**: Statistics, Ecology, Mathematical statistics

## Introduction

The statistical ecology develops rapidly with the advancement of analyzing technology and method. Chinese scientists and ecologists made great efforts editing the ecological writings such as "Environmental statistics" (Wu 1991), "An introduction to energy ecology" (Zu 1990) etc.

The development of statistical ecology experienced the following three main stages:

(1) Quantitative measurement, classification by table, direct gradient analysis and common statistical analysis (Greig-Smith) etc..

(2) Linear indirect gradient analysis, linear PCA ordination based on plant versus environment. Polar ordination with any projection axis. RA reciprocal averaging. CA correspondence analysis, canonical variance, CCOA canonical correlation analysis based on environment factors versus plant.

(3) Utilization of Computer. Flexible Shortest Path Adjustment. Detrended Correspondence Analysis. Comprehensive correspondence analysis and some other classification and ordination.

There are many analysis softwares (parcel), such as MULVA, SYNTAX, PATN, CEP etc. In China, Yang Hanxi (1988), Gao Qiong (1991), Zhong Yang (1990) had done some researches about software. FUEPAK quantitative analysis parcel (by Gao Qiong) is made of DPCOA and DCPCOA ordination to solve models cluster problems with information theory.

## Data collection

Sampling is important, which means how to decide the Sampling Unit (SUs).

The successful sampling units used in ecology including quadrates, leaves, light traps, soil cores, pit traps, individual organisms, and belt transects. Some occur naturally (e. g., leaves), others are arbitrarily defined (e. g., quadrates). The data collected from measurements are tabulated into an ecological data matrix, which depended on the research purpose.

The researches are both about temporal dynamics (e. g., community succession) and spatial distribution. The stages of an observational ecology approach are as followings:

(1) Data collection (including research scope, sampling, measurement, data matrix).

(2) Data analysis (including species in sampling units, similarity measure, structure, biotic patterns).

(3) Description and interpretation

(4) Prediction and hypotheses

(5) Models and checking up.

There is the problem about the investigation of the forest structure (or species diversity), that means how much the site is, including most of the plant species. China's forest areas vary from subtropical to cold-temperate zones including some monsoon forests. Qu Zhongxiang thinks there are more plant species in a more complicated community possessing better environmental condition. The minimum area is no less than 2 500 $m^2$ in tropical rainforest (in Xishuangbanna), which is only 400 $m^2$ in Korean pine community in northeastern China. Yang Chi (applying Greig-Smith pattern analysis theory) thinks the area of 500 $m^2$ is the optimum area for grassland distribution pattern analysis.

## Spatial and time patterns analysis

Ecologists consider the importance of spatial patterns in communities and identified various casual factors that may lead to patterning of organisms. It is made by the biology distribution in the environment and their relationships. The community time pattern is the character of its dynamics including the species

**Received**: 1999-07-14
Responsible editor: Zhu Hong

periodical changes in time structure due to the time rhythm of natural factors and the community change from one type to another type in its long developing process. There are three basic types of patterns: random, clumped and uniform.

Wang Bosun (1987) analyzed the dominant forestry population qualitative dynamics in Dinghushan. The result showed that the growth curve of *Cryptocarya chinensis* and *C. concinna* conformed to logistic curve. Zu Yuangang and Zhang Xijun (1994) used CTM (Continuous Time Marcof) as modeling that means to set up the harmonious developing model of human population resources and environment in Songnen plain by the mathematical type composed of human population, resources and environment. It provided the evidence to improve the environment and develop the economy. Fang Jiming and Sun Ruyong (1991) studied about the seasonal change of Bandt's voles spatial distribution, concluded that it was random in initial stages of its reproduction, and clumped in other stages. They also analyzed the spatial competitive relationship between male and female rats. The population trends to be clumped in reproductive period, trends to diffuse from nonproductive to reproductive period then to be clumped after reproduction. Zu Yuangang and Ma Keming applied Box dimension of fractal theory and semivariance dimension combined of fractal theory and geostatistics to study the '*Aneurole pidium chinensis* horizontal' distribution patterns on the population and community scale in Songnen plain. The results showed that the population horizontal distribution patterns were self-similar and *Aneurole pidium chinensis* population was always the dominant population. Soil is the essential factor to affect the community distribution patterns. Zu Yuangang (1990) studied about the dynamic changes of energy accumulation rate of *Aneurole pidium chinensis*, they are as follows: (1) the daily changes, the seasonal changes, the seasonal changes of relative and absolutely accumulative rates; (2) the daily changes of energy loss rates in community evapo-transpiration, the seasonal changes the seasonal changes of relative and absolutely accumulative rates; (3) the daily changes of energy loss rates in community dark respiration, the seasonal changes the seasonal changes of relative and absolutely energy loss in community dark respiration etc.

## Species-abundance and species-area relations

One of the most obvious and consistent phenomena observed in ecological communities is the variation in species abundance as well as species areas. For example, in a given community, how many species are there and what is their relative abundance? How many species are rare? How many are common? The mathematical descriptions have led to numerous theories regarding community stability, energy pathways, resource partitioning, species-area relationships and evolutionary process.

There are two good models about species-area:

Kylin (1926) model:

$$S_q = S_\infty (1 - e^{-mq})$$

Where $q$ is investigated area; $S_\infty$ is the total species number, $S_q$ is the present species number, $m$ is a nondimension parameter.

Fisher (1934) model:

$$S_q = a \bullet \ln(1 + \frac{dq}{a})$$

Where $d$ is the density per area (including all species).

There are $\alpha$ diversity, $\beta$ diversity and $\gamma$ diversity, also many indices are introduced. We will give the diversity number proposed by Hill (1973). They are very popular in China's ecological papers.

The equation is:

$$N_A = \sum_{i=1}^{r} (p_i)^{\frac{1}{(1-A)}}$$

Where

$N_0 = s$ is the total number of the species; $N_1 = e^{H'}$ is the number of abundant species in the sample;

$N_2 = 1/\lambda$ is the number of very abundant species in the sample

where $H'$ is Sannon's index, $\lambda$ is Simpson's index.

$$H' = -\sum_{i=1}^{H} p_i \ln p_i$$

$$\lambda = \sum_{i=1}^{H} P_i^2$$

Where $P_i$ is the number proportion of species $i$ to the total number of individuals in community.

Diversity index includes species richness and evenness. The community diversity presents in space and time.

Evenness indices

Hurbert (1971) noted the evenness indices either as

$$V' = \frac{D}{D_{max}}$$

or

$$V = \frac{D - D_{\min}}{D_{\max} - D_{\min}}$$

Where $D$ is observed diversity index, $D_{\min}$ and $D_{\max}$ are the minimum and maximum values respectively, The most common evenness index used by ecologists is:

$$E_1 = \frac{H'}{\ln(s)} = \frac{\ln(N_1)}{\ln(N_0)}$$

When all species in a sample are equally abundant, the evenness index should be maximum and decrease to zero as the relative abundance of the species diverge away from evenness.

Richness indices

Two historically well-known richness indices are as follows:

$$R_1 = \frac{s-1}{\ln(n)}$$

$$R_2 = \frac{s}{\sqrt{n}}$$

Where $S$ is the number of species and $n$ is the number of the total individuals.

There are many plants and animal species in low latitude tropical habitat, and the diversity is much higher than that in temperate and Arctic zones. Diversity increases from high latitude and high altitude areas to low latitude and low altitude tropical rainforest and coral reefs. Zhang Rongzu found that the mammal species in China and its neighbor areas are related to the latitude, average temperature, rainfall in inland arid areas and the altitude. The variation is also obvious ($r$=0.949, $P<0.05$).

The community stability includes resistance and resilience, and it is due to the reaction ability of population and individuals to the environment changes or disturbance. When there is more species in a community and they distribute evenly, the community will have a stronger feedback system and buffer the disturbance better. Many ecologists accepted that diversity results into stability, but someone doubted about it and there are still some quarrels about it. Papers, journals and books about diversity researches are booming in the late years. "China's biodiversity" and "21 century agenda" are published.

## Population dynamics

The *Ruditapes philippinarum* population dynamics and reproduction protection were studied in Ports Xiangshan and Huangdunzhi.

$$l_i = \frac{4.06}{1 + e^{2.1792 - 0.0861t}{}_m}$$

They proposed the biologic index (style, age), fishing limits and season of the important economic shell.

Zu Yuangang and Chen Jihong (1993) concluded that in Songnen grassland the nine communities aboveground biomass and their dominant population aboveground biomass growth confirm to logistic curve and they are related to the rainfall and temperature in growing season. Zu Yuangang and Yang Jingchun (1989) applied the three-dimensioned spatial trend face analysis to reveal the comprehensive effects of northeastern *Aneurolepidium chinensis* grassland soil temperature and water-containing capacity on the respiration rate of soil microbiotics. The trend is that the respiration rate of soil macrobiotics increases with the soil temperature and water-containing capacity, the metabolism also becomes stronger.

## Species affinity

Ecological community consists of many coexisting species. Their competition on the common resources is interspecies affinity degree, represents as niche overlap, which is how to use the common resources. Interspecies association is that two species choose or avoid the same habitat, attract, exclude or no relative. Co-variation is the reaction way of environmental factors while the abundance trend to increase or decrease.

Dong Shilin, Zu Yuangang and Liu Xinghua (1992) applied niche breadth measurement Smith equation to compute the niche breadth and overlap of 10 major plant Populations on one-dimension environment factors in *Aneurole pidium chinensis* grassland. They also calculated the niche breadth on the gradients of two-dimension and three-dimension environmental factors.

1. The niche breadth of species $i$:

$$B_i = \frac{1}{\sum_j^r P_{ij}^2}$$

2. Hurbert (1978) suggested the overlap index between species 1 and species 2 ($HO$):

$$HO_{1,2} = \sum_j^r \frac{P_{1j} P_{2j}}{C_j}$$

Where $C_j$ is the relative abundance of the $j$th re-

source.

3. Interspecies association Jaccard Index (*IJ*):

$$IJ = \frac{a}{a+b+c}$$

Where *a* is the number of SUs where both species occur; *b* is the number of SUs where species A occurs, but not species B; *c* is the number of SUs where species B occurs, but not species A.

Wu Jincai and Pang Xiongfei (1990) applied composite design of quadratic regression rotation to study about the preying of four species of predator spiders (*Pirata subpiraticus*, *Oedotherax insecticesps, Clubiona japonicola*, *Bianor hotingchiehi*) and brown plant hopper/nymph (*Nilapavata lugens*) coexisting mathematical models and to forecast the quantitative changes of *Nilaparata lugens*. The result shows that there exist disturbance and killing among the species in a complicated agri-ecosystem.

The classical Predator-prey calculus equation assumed that when the population was alone, the prey increased in exponent and the predator decreased in exponent. Their consistent generation mathematical model is $dH/dt=a_1 H$; where $a_1$ is the inner growth rate when the population is alone, *H* is population density of prey, *t* is time.

Zhou Jizhong and Chen Changming (1987) first proposed the mathematical model of quantitative measurement about the switching of predators to prey, they also studied about the choosing preying of *Lycosa pseudoannulata* on *Nilaparvata lugens* and *Cnaphalocrocis medinalis*, and suggested the model of one predator population on two prey populations.

## Community similarity and community cluster analysis

There are two indices of community similarity. One is the real similarity index, which reflects the similar degree among the entities. When the two entities are completely the same, the index is maximum and it is minimum while they are completely different. The other one is heterogeneous index, it reflects the heater degree among the entities, the bigger the value is, the less the similarity is, and the less the value is, the bigger the similarity is.

The matching coefficients are important similarity index, which is designed by the matching relationship of the occurrence of some characters in the entity (or sample).

$S_1=(a+d)/(a+b+c+d)$

Where *a* is the number of species that occur in both samples; *b* is the number of species that occur in sample A but not sample B; *c* is the number of species that occur in sample B but not sample A; *d* is the number of species that don't occur in both samples.

Cluster analysis (CA) is a classification technique for placing similar entities or objects into groups or clusters so that there is the biggest similarity in one group and the biggest difference among groups.

(1) The nearest-neighbor or single linkage is that the nearest distance of the two samples is chosen as the distance of the entities.

(2) The furthest-neighbor or complete linkage is that the furthest distance is chosen as the distance of the entities.

(3) Heterogeneous coefficient: distance coefficient Euclidean distance is common:

$$D_{1(j,k)} = \sqrt{\sum_{i=1}^{p} (x_{ij} - x_{jk})^2}$$

or mean Euclidean distance:

$$D^2_{2(i,k)} = \frac{1}{p}\sum_{i=1}^{p} (x_{ij} - x_{ik})^2$$

Where $x_{ij}$, $x_{jk}$ are species *j* and *k* in sample *i*.

Others are generalized distance and absolute value distance etc. Guo Yiquan and Zhao Zhimo (1992) proposed food net similarity index and its measurement based on food net reflect matrix. The measurement results of three orange and tangerines gardens *Coccoidea*-parasite fly similarities are that: The biggest similarity is between biological prevent garden and comprehensive prevent garden, the bigger similarity is between chemical prevent garden and biological prevent garden, and the least is between comprehensive and chemical prevent garden.

Zhang Jintun (1985) applied Fuzzy cluster analysized to *Vitex negundo* var. *heterophlla* shrubs in Shanxi Province and compared the results with those from group mean analysis and graphic theory cluster analysis. He concluded that the Fuzzy cluster analysis had obvious ecological meanings. Zhou Xinyuan simplified population dynamics of *Panonyehus citri* by Fuzzy cluster analysis into its population dynamics.

Some others as information clustering analysis, probabilistic clustering analysis, dynamic clustering analysis and optimum section analysis are used in ecological researches.

## Community ordination

Ordination is the techniques in which entities or sampling units (SUs) are arranged in relation to one or more coordinate axes such as their relative positions to the axes and to each other provides maximum information about their ecological similarities. In

another word it means the spatial arrange of entities.

**Polar ordination (PO)**

It is founded in 1950's. The other name is Bray-curtis ordination. Jiang Youxu (1982) studied about the ordination of subalpine forest vegetation in western Sichuan Province. First he classified the local communities into 5 association cycles including 32 associations and applied the data from 6 typical associations to ordinate. The species included 12 arbors, 53 shrubs, 115 herbs and 55 bryophytes.

**Principal components analysis (PCA)**

The ecological ordination regards to numerous biological and environmental factors. Some minority factors can be extracted from the complicated factors to substitute for the whole information so that the result from minority factors' information is similar with that from all the factors' information. PCA was first used in ecology by Goodall. It's often referred to as one of the eight analysis ordination methods. It also can be used into Euclidean distance matrix.

Zhao Zhimo and Guo Yiquan (1985) applied PCA in the orange and tangerines gardens in Chongqing. They studied about the law of insects population varies with the ages of trees in 10 different age class gardens. The results show that the lst, 2nd and 3rd principal components respectively keep 33.92%, 23.45%, and 15.82% information of the multiple-dimension space. The biggest contribution to the 1st principal components is from *Aleurocanthus spiniferus* and *Nosema tortricis*. Their burdens are 0.923 1 and 0.833 8 respectively. They are all the pests injuring the new leaves. The biggest contribution to the 2nd principal component is from *Creambycidae* and *Prontaspis yanonesis*. They injured trees. The biggest contribution to the 3rd principal component is *Eotetranychus kankitus* and *Panonychus citri*. Ren Shiping (1986) applied PO and PCA to study about the quantitative classification of tropical grassland in Hainan Island; He classified the 19 samples into 3 types, 9 communities. The distribution pattern is closely related to the slope, grazing degree and soil fertilization, some others as Principal axes analysis (PAA), correspondence analysis, reciprocal averaging, discrimination and canonical correlation analysis etc. The canonical correlation analysis is usually used in psychology and pedagogy.

## References


Cao Guangxia. 1991. Life table of *Picea purpurea* and *Abies jaxoniana*. Acta ecologica sinica, **11**(3)

Chen Jihong. 1993. The study on the law of aboveground biomass growth of the main plant communities in Songnen grassland. (Master Theses) . Harbin: Northeast Forestry University

Greig-Smith, P. 1983. Quantitative plant ecology. Oxford London Edinburgh, Boston Melbourne

Guo Yiquan and Zhao Zhimo. 1992. The measurement of the similarity among the communities food nets, **11**(3) Chinese Journal of ecology

Liu Xinghua. 1992. Preliminary study on the niche of plant population in *Leymus chinensis* steppe. (Master Theses). Harbin: Northeast Forestry University

Ludwing, J.A. and Reynolds, J.F. 1988. Statistical Ecology-a primer on methods and computing. Awiley-interscience Publication John Wiley and Sons

Ma Keming. 1994. A FRACTAL study on horizontal distribution pattern of *Aneurolepldium chinensis* community. (Master Theses) . Harbin: Northeast Forestry University

Pielou, E.C. 1991. Mathematical Ecology. Science Press

Southwood, T.R.C. 1984. Ecological Methods with particular reference to the study of insect populations. Science Press

Su Xiangyao and Lin Changshan. 1986. The study on the model of Mythimna separates population dynamics. Acta ecologica sinica, **6**(1)

Wan Bosun. 1987. The dynamics of the dominant populations in Ding hu-shan. Acta ecologica sinica, **7**(3)

Wang Yihong. 1990. The experimental methods of forest ecology. Harbin: Northeast Forestry University Press.

Whittaker, R.H. 1986. Ordination of plant communities, Science Press

Whittaker, R.H. 1985. Classification of plant communities, Science Press

Wu Jincai and Pang Xiongfei. 1990. The Mathematical models of multispecies population preying and its forecast in *Nilaparvata lugens*. Acta ecologica Sinica, **10**(3)

Wu Yuming. 1991. Environmental statistics. China Environmental Science Press

Xia Naibing. 1992. The analysis of the diffuse trends of Chinese pine caterpillar larvae in its tree climbing time sequences. Acta ecologica sinica, **12**(1)

Zhou Jizhong. 1987. The preying action of pirate subpiratious to *Niaparvata lugens* and its model research. Acta ecologica sinica, **7**(4)

Zhang Xijun.1994. A CTM model for coordinate development of population, resources and environment in Songnen Plain. ( Doctoral Dissertation ) . Harbin: Northeast Forestry University

Zhao Zhimo and Guo Yiquan. 1991. The theory and method of community ecology. Chongqing: Chongqing Science and Technology Press

Zu Yuangang. 1990. Introduction to Energy Ecology. Changcun: Jilin Science and Technology Press